



TITLE:

A Study on Fine-Grained User Behavior Analysis in Web Search(Abstract_要旨)

AUTHOR(S):

Umemoto, Kazutoshi

CITATION:

Umemoto, Kazutoshi. A Study on Fine-Grained User Behavior Analysis in Web Search. 京都大学, 2016, 博士(情報学)

ISSUE DATE:

2016-03-23

URL:

<https://doi.org/10.14989/doctor.k19852>

RIGHT:

学位規則第9条第2項により要約公開; 許諾条件により要約は2016-04-09に公開; 許諾条件により本文は2018-11-13に公開

(続紙 1)

京都大学	博士（ 情報学 ）	氏名	梅本 和俊
論文題目	A Study on Fine-Grained User Behavior Analysis in Web Search (Web検索における細粒度ユーザ行動の分析に関する研究)		
(論文内容の要旨)			
本論文では、Web検索システムの検索性能の向上を目的として、粒度の細かいユーザ行動の分析に基づく検索意図推定手法が提案されている。			
Web上のコンテンツの充実にともない、人々はさまざまな目的のためにWeb検索エンジンを利用するようになってきている。情報検索に関する研究分野では、適切な検索結果を提示することを目的として、ユーザの検索行動の分析に関する研究がこれまでに数多く行われてきた。ユーザがWeb検索エンジンを用いて行うタスクの中には、1つのクエリの投入や1つの文書の閲覧では達成することの困難な複雑なものが存在する。複雑な検索タスクには、単純なタスクと異なる特徴として、いくつかの共通点があげられる。最初の2つは、(1) 同じクエリが投入されたとしても、検索者や検索コンテキストに応じて適合情報が異なる、および(2) ユーザがタスク中に検索する内容は時間とともに変化する、という検索意図に関する特徴である。また、複数文書の収集が必要であるという要件にともない、(3) 単一の適合文書の閲覧が必ずしも検索タスクの成功を意味しない、および(4) 検索をいつ終了すべきかの判断はユーザにとって容易ではない、という検索成果に関する特徴も存在する。従来の検索行動分析に関する研究の多くは、多くのユーザにとって代表的な意図の発見を目的としており、クエリ単位や文書単位といった粒度の粗いデータに焦点が当てられてきた。そのため、こうした複雑なタスクにおける、個々のユーザの検索に関する特徴は考慮されてこなかった。			
本論文では、こうした背景を踏まえ、複雑なタスクにおける個々のユーザの検索行動の分析および活用に関する提案を行っている。本論文の核となるアイデアは、複雑なタスクにおけるユーザ／文書間の細粒度の検索インタラクションデータを用いるという点である。学位申請者は、具体的には、検索意図の理解に関する課題として「注目行動データを用いた興味語の抽出」および「検索行動データを用いたクエリ修正の予測」に関する研究を、検索成果の評価に関する課題として「検索情報の不一致が検索行動に与える影響の分析」および「検索結果中の未閲覧情報量の可視化」に関する研究を行い、提案手法の有効性を評価している。			
本論文は全7章から構成されている。その概要は以下の通りである。			
第1章は序論であり、本論文の研究の背景、本論文の研究を行うに至った動機、ならびに本論文の研究の全体の概要を述べている。			
第2章では、第3章から第6章において説明される研究課題に関して、それらに深く関連する分野の従来研究を整理し、本研究との位置付けを議論している。			
第3章では、クエリの背後に隠れた検索者の意図に関連する語集合を検索者の行動データから推定する手法を提案している。学位申請者は、その推定にあたり、Webページ閲覧時の検索者の視線の動きに着目している。提案手法は、「検索者に多く注目された語は検索意図に関連する可能性が高い」という仮定に基づき、検索者の視線データから注目語を抽出し、全閲覧ページ集合に対してその集約を行う。提案手法の有効性は、ユーザ実験によって評価されている。評価実験の結果、提案手法は個々のユ			

ユーザの意図に即する検索に適切な語集合を推定可能であることが示されている。

第4章では、タスク中に変化するユーザの検索対象を事前に推定する手法を提案している。学位申請者は、本タスクを、現在の検索に関する情報が与えられた際に、次の検索に用いられるクエリの修正タイプを、既存の分類である5種類の中から予測する問題と形式化している。提案手法は、現在の検索時におけるさまざまな粒度の行動データを用いて分類器を構築し、次の検索クエリの修正タイプを予測する。学位申請者は、提案手法の分類精度の評価に加えて、個々のクエリ修正に先立つユーザ行動に関する詳細な分析を行っている

第5章では、検索情報の不一致がユーザの検索行動および最終的な検索成果にどのような影響を与えるかを分析している。学位申請者は、その分析に際して、情報検索に関するユーザの専門知識（検索専門性）に着目している。具体的には、事実発見型タスクに関するユーザ実験において被験者が発見した答えを抽出することでデータセット作成し、それに対して、（1）検索情報の不一致がユーザの最終的な満足度を与える影響、（2）検索情報の不一致に気付いたユーザが検索を終了するタイミング、および（3）時間の経過にともなう検索対象の情報の変化という3種類の観点から、分析が行われている。学位申請者は、分析によって得られた知見に基づき、検索専門性の低いユーザを支援するための枠組みについて議論している。

第6章では、トピックに関する重要な情報の網羅的な収集が望まれるタスク（網羅的情報収集タスク）における未探索の観点の発見や検索終了判断の支援を目的として、検索結果に含まれる未閲覧情報量を提示するクエリ推薦インタフェースを提案している。学位申請者は、未閲覧情報量を検索トピックに関する重要な情報が検索結果の中に残っている度合いと定義し、検索結果中の未閲覧文書集合から得られる追加利得としてその値を推定している。また、文書集合に対する利得は、（1）観点の重要性、（2）観点の新規性、および（3）観点に関する文書の適合性に基づき定式化されている。提案インタフェースの有効性を評価するために、24名の被験者を対象に4種類の検索トピックを用いたユーザ実験が実施されている。その結果、提案インタフェースの利用者は、未閲覧情報量の推定精度が高いトピックにおいて、網羅的情報収集タスクを効率的に達成可能なことが確認されている。

第7章では、本研究で得られた研究成果をまとめ、さらに今後の展開について議論している。

注) 論文内容の要旨と論文審査の結果の要旨は1頁を38字×36行で作成し、合わせて、3,000字を標準とすること。

論文内容の要旨を英語で記入する場合は、400～1,100 wordsで作成し
審査結果の要旨は日本語500～2,000字程度で作成すること。

(論文審査の結果の要旨)

本論文は、Web検索システムの検索性能の向上を目的として、複雑な検索タスクにおける粒度の細かいユーザ行動の分析に基づいて検索意図の推定などを行う手法について提案を行い、その有効性を検証したものである。

適合性の高い情報を検索することは情報検索システムの本質的で重要な課題である。そのため、情報検索分野では、ユーザの検索行動を分析することで、クエリの背後に隠れた検索意図のマイニングやクリックされた文書の適合性の推定を行う研究が盛んに行われてきた。しかし、これらの研究の多くは、クエリ単位や文書単位といった粗い粒度のユーザ行動を分析したものであり、クエリ投入や文書閲覧が複数回行われる複雑なタスクについては、これまであまり焦点があてられてこなかった。多目的かつ多様な意図の下で検索エンジンが使われるようになっている現代において、複雑な検索タスクにおけるユーザ行動の理解は重要な研究課題の1つである。

学位申請者は、こうした複雑なタスクを対象として、ユーザ／文書間の細粒度の検索インタラクションデータに着目し、個々のユーザの検索行動を分析し検索意図を推定する手法を提案している。さらに、実際のユーザをとまなう実験によって、提案手法の有効性を評価するとともに、興味深い結果を示している。

具体的には、本論文の研究によって得られた成果は以下のように要約される。

1. 検索を通じて適合情報が明確化されるタスクにおいて、Webページ閲覧時の検索者の視線データから、検索者の意図に適合する語集合を推定する手法を提案した。具体的には、「検索者に多く注目された語は検索意図に関連する可能性が高い」という仮定を置き、ページ中の語の出現回数が検索者の注目行動に与える影響を考慮した複数の推定モデルを提案した。ユーザ実験によって、推定精度の高いモデルを明らかにするとともに、提案手法が検索に適切な語を推定可能であるという興味深い結果を得ている。
2. 複数回のクエリ投入が必要なタスクにおいて、次の検索クエリの修正タイプを事前に予測する手法を提案している。提案手法は、現在の検索に関するさまざまなユーザ行動データを用いてSVM分類器を構築することで、その予測を行う。クエリや文書だけでなく、視線のような細粒度データが、クエリ修正タイプの予測に大きく貢献することを、実験によって示している。
3. 検索情報に不一致が存在するタスクにおいて、情報検索に関する専門知識の有無（検索専門性）がユーザの検索行動と検索成果に与える影響を分析している。本分析によって、(1) 検索専門性の高いユーザの満足度と検索情報の不一致との間に負の相関関係が存在する、(2) 検索専門性の高いユーザは、答えの発見以後も長い時間をかけてタスクに取り組む、ならびに(3) 検索専門性の高いユーザは、タスク開始から一定時間が経過後も、検索情報の絞り込みを行わない、といった興味深い知見を得ている。
4. 情報の網羅的な収集が望まれるタスクにおけるユーザ支援を目的として、検索結果に含まれる未閲覧情報量を提示するクエリ推薦インタフェースを提案している。学位申請者は、クエリに対する未閲覧情報量を検索結果中の未閲覧文書集合から獲得可能な追加利得として定式化し、サブトピックマイニング手法を利用することでその値の推定を実現している。提案インタフェースの利用者は、未閲覧情報量の推定精度の高いトピックにおいて、(1) 重要な情報を十分に調べた後で個々の検索を終了する、(2) 重要な情報を多く含むクエリを次の検索に利用する、(3) タスクの後半にも重要な情報を効率的に収集する、およ

び(4)単位時間あたりに獲得可能な利得が上昇する、といったことがユーザ実験の結果から示されている。

以上のとおり、本論文では、Web検索システムの検索性能の向上のために、複雑な検索タスクにおけるユーザ／文書間の細粒度の検索インタラクションデータの分析に基づく検索意図推定手法を提案している。Web上での多様なコンテンツの登場やサービスの発展にともない、検索エンジンが単なる情報収集ツールとしてだけでなく、より高度な意思決定のために使われるようになっていくことを鑑みると、本論文が対象としている複雑な検索タスクは、今後の情報検索に関する研究において重要な位置を占めると考えられる。また、学位申請者が提案した細粒度ユーザ行動の分析手法は、従来の粗い粒度の検索行動の分析に関する研究と比較しても、新規性が高いと言える。このように、本論文は、学術上寄与するところが少なくないため、博士(情報学)の学位論文として価値あるものと認める。また、平成28年2月19日に論文内容とそれに関連した事項について試問を行った結果、合格と認めた。

注)論文審査の結果の要旨の結句には、学位論文の審査についての認定を明記すること。
更に、試問の結果の要旨(例えば「平成 年 月 日論文内容とそれに関連した口頭試問を行った結果合格と認めた。」)を付け加えること。

Webでの即日公開を希望しない場合は、以下に公開可能とする日付を記入すること。
要旨公開可能日： 年 月 日以降